What does it take to make a first word? The development of speech motor control during the first year of life

Claire Lalevée, Anne Vilain

Institut de la Communication Parlée, UMR CNRS 5009 - Université Stendhal, DU, 1180 Av Centrale, BP25 - 38040 GRENOBLE Cedex 9, France

claire.lalevee@icp.inpg.fr, anne.vilain@icp.inpg.fr

Abstract – *All babies in the world utter their first words around the age of 12 months, i.e. first productions identified by the parents as lexical items of their mother tongue. Studies have shown that the nature of the segmental inventory is governed by biomechanical constraints bearing on the developing supra-glottal motor system. The features that make a first word are then to be looked for in the prosodic shape of the word: laryngeal and rhythmic control, and in what we called "closance" control: the control of pressures inside the vocal tract. We are examining these developing features on the basis of longitudinal audio-visual recordings of a French child from the age of 6 to 15 months.*

1. Introduction

The first year of life is a crucial period for speech development in infants. Around 6 months is the time when babbling appears, under a form which is quite similar for all children in the world, whatever the language in which they are reared. It is a period when the child has no control over the nature of his productions and no ability to produce phonological units of his mother tongue. Around 12 months the child begins producing his first words i.e. his first meaningful utterances. What has been taking place during this period is the emergence of phonology, which is to be related to the emergence of new controls over the child's vocal apparatus. Several studies have hypothesized (MacNeilage, 1998, Munhall, 1998, Green, 2000), and we have brought support for this view (Brosda, 1998, Matyear et al., 1998, Davis et al., 2000, Vilain et al., 1999), that canonical babbling, a major step in language development, can be related to the sudden emergence of control over the carrier articulator, i.e. the mandible, in the absence of any other voluntary control over the other speech articulators such as the lips and the tongue. Babbling utterances are thus byproducts of the oscillations that constitute proto-syllabic frames, and intra-syllabic variegation will appear late after the emergence of babbling and increase with the progressive mastering of the carried articulators. No major change in the segmental inventory of the child happens between 6 and 12 months, and that the first words he produces share the same inventory as babbling, with "mama", "papa", "tata" as the most frequent first words (Vihman, 1986). What has been evolving during that period, and what brings a child from nonsense "mama" to "mama" as a word of his mother tongue is what we called *closance* control, a better control of the pressures and coordinations in the vocal tract, from the glottis to the lips, that will enable them to become true adult-like syllables. Controlling the pressures and impedances in the resonators of the vocal tract will produce better inter-segmental contrasts, and this is made possible in particular through two types of controls, that we will discuss here: (i) the control of the velum, which yields a fully oral vocal tract to produce salient consonant-vowel sequences, and (ii) the control of the oro-laryngeal coordination to sharpen the distinction between close and open configurations, i.e. closants and vocants. And a third kind of control can be added to make words out of protosyllabic frames without any improvement in supra-glottal articulatory control: (iii) the rhythmic mandibular control, which enables the child to adapt to the prosodic patterns of his mother tongue. We have investigated the evolution of these features in a corpus (Brosda, 1999) consisting in audio-visual recordings of a French child aged from 6 to 15 months. The infant has been recorded in his family environment during 45 minutes every fortnight, from the date when the parents signalled the beginning of babbling. The vocalizations that were retained are productions that corresponded to the babbling criteria (Koopmans Van Beinum, 1999).

2. Closance control: development of the prosodic word

2.1. Development of the oral/nasal control.

A close look at the oral/nasal control?

One of the major features that the baby has to acquire for speech is the oral/nasal control, in order to produce sequences with salient auditory contrast between closants and vocants, as it is the case in adult speech syllables. Nasal consonants represent 14.6% of the consonantal sounds of the world's languages (Stefanuto, 1999), nasal vowels are 22.4% of the vowels (Vallée, 1994). A sequence consisting of a nasal consonant followed by a vowel has little auditory contrast, being constituted of two loud elements, since the nasal consonant, as well as the vowel, shows a large amplitude and a formant structure. Such a sequence hardly shows the non-sonorant – sonorant pattern that characterizes as a majority a syllable in the world's languages. Moreover nasals' places of articulation are poorly identifiable for the listener: languages do not make use of many different places of articulation. A commonly-held view is that oral sounds are produced with an active raising movement of the velum against the pharyngeal wall, and that it is a lowering movement of the velum which causes nasal sounds. A number of studies (Kinney & MacNeilage, 1997) have shown that baby vocalizations contain a large proportion of nasal sounds. The question for developmental studies is then: what is the default position of the velum from birth? How and when is the control of this particular velum gesture acquired by the baby? But the dichotomy between nasal and oral sounds has to be considered in closer detail in order to better understand speech control development and phonology.

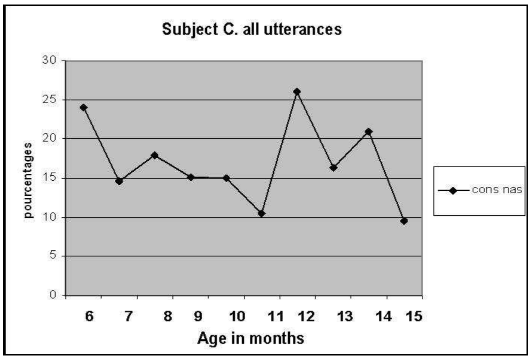
First a simple articulatori-acoustic statement about nasal consonants and vowels: opening the nasal tract by way of lowering the velum has very different aero-acoustic consequences, whether the oral tract is open or closed. If the oral tract is closed as it happens in occlusives, the nasal fossae will act as a resonator and induce the appearance of formants, therefore causing a dramatic change in the acoustic structure of the consonant. Thus the difference between the two positions of the velum in a consonant yields very obvious and unambiguous cues on the speech signal: from silence to a formant structure. By contrast, starting from an open vocal tract to lower the velum, i.e. changing from an oral to a nasal vowel does not modify radically the acoustic nature of the sound, but has rather slighter effects, such as widening the bandwidth of the first formant, lessening the amplitude of this first formant, lowering the frequency of the second formant (Delattre, 1954). Actually, the detection of reliable acoustic correlates of nasality in vowels still represents a stumbling block for acoustic researches. Spectral slope for example was once supposed to be a good correlate of nasality, but it proved unreliable for automatic detection (Kinney, 2001). Detecting nasal consonants rather than nasal vowels is therefore more reliable, and all the more in baby vocalizations.

A second point in our argument is given by an articulatory study by Rossato et al. (2003) and Amelot & Rossato (2006). They have studied the movements of the velum in a French speaker through electro-magnetic articulometry measurements, and particularly the distribution of velum vertical positions for oral versus nasal vowels, and oral versus nasal consonants. Nasal vowels are here to be understood as parts of the French phonological nasal category, and acoustically analyzed as nasal. What comes out of this study is that the difference from an oral to a nasal consonant can be realized by a mean 2 mm movement of the velum, whereas a nasal vowel needs a much more widely open nasal tract and so a larger downward movement of the velum (mean difference between the high and low positions of the velum: 7 mm). It is to be related to the impedance between the nasal tract and the oral tract: the impedance of the nasal resonator remains too high if the nasal port is not open enough while the oral port is open. This implies that changing from the oral to the nasal mode in consonants represents a small effort, needing little energy, while the change in vowels is an effortful gesture. Moreover it appears that an oral vowel can even be produced with a small aperture of the nasal tract, especially for low vowels. In other words, a [mama] sequence can be produced with a stable slightly lowered position of the velum. Getting back to development, these results mean that while the slightly open configuration of nasal consonants may be considered as a default position of the velum in the early stages of development, the low position for nasal vowels will imply a complex control of the velum that should develop much later in ontogeny. This sheds light on the studies that should be carried out about nasality in babbling vocalizations: no strong effect is to be expected in the acoustic structure of vowels until the time when the baby is able to control a large opening of the velopharyngeal port, while nasal consonants may show the evolution of a first stage in the development of this control.

Developmental hypothesis: nasal consonants first, then vowels ?

Our hypothesis for baby speech is that the development of the oral / nasal control unfolds in 3 successive stages. The first stage would be when babies start with the default position of the velum and produce as a majority nasal consonants and oral vowels. The second stage would be when the baby gains the first control of the raising of the velum for oral consonants, and the third stage would be a second control that needs more energy and will take place later: the lowering of the velum for nasal vowels. We have started by studying the evolution of the proportion of nasal consonants in babbling, to determine when the baby reaches adult-like norms. The first outcome of the study is the number of nasal consonant-like sounds, as a proportion to oral occlusive consonants for the two children. We are not comparing nasals with consonants other than occlusives, since nasality would not be detected in a non-occlusive consonant. The first observation we can make (cf. Figure 1) is that the proportion of nasals keeps lessening from seven months on. At 11 months, the babies have reached the proportion of nasals in French, their native language (10, 2%: Tubach, 1990), a fact that we may consider as an evidence for a first control to raise the velum. It is to be noted that by the age of 12 months, the proportions of nasal sounds suddenly increase greatly in the infant’s productions. We found out that this phenomenon is to be explained by a lexical effect, with the appearance of words such as “non” (no) and “maman” (mummy) in the children's language. When the words « non » and « maman » are removed, the proportions of nasal consonants decrease to reach French adult proportions from 11 months to15 months (cf. Figure 2).

What becomes clear from our study is the difference that has to be made between nasality in consonants and in vowels, which should actually be considered as two different features with two different types of control. The recordings in our corpus do not show nasal vowels at all; even the vowels that are produced in a nasal context are not nasal when closely considered: e.g. "maman" is not produced as [mamã] but as [me:na]. The mastering of the low velum position which is necessary for nasal vowel production is to be tracked later on in speech development.



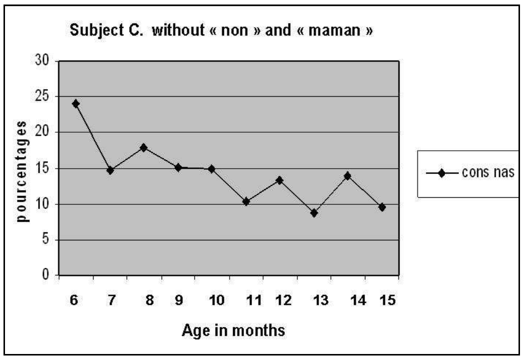
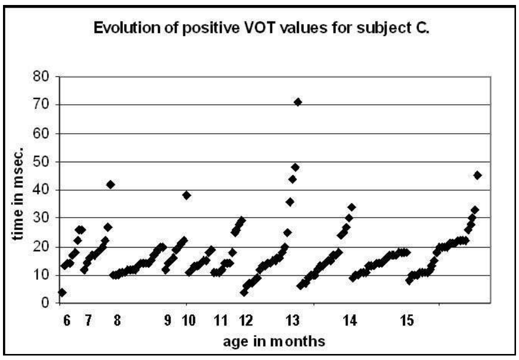


Figure 1 (left): Evolution of the nasal consonants vs. oral consonants proportions, 6 to 15 months.
Figure 2 (right): Same evolution without the first words "non" and "maman"

2.2. The oro-laryngeal control.

The oro-laryngeal coordination, i.e. the coordination between the glottal and supra-glottal gestures, is a basic feature in human languages. The voiced/unvoiced distinction has not been much studied in babbling even though this distinction is present in a very important majority of the world's languages. Studies (Eimas, 1971) led on this coordination focus on the perception of the VOT (Voice Onset Time) defined as the time interval between the beginning of the periodic glottal pulses and the release of the supra-glottal occlusion (Lisker & Abramson, 1964). So we have studied VOT production and evolution from our corpus. VOT measurements were made on 569 bilabial, coronal and dorsal occlusives in initial position in CV context. Results show the same duration for positive VOT values from 6 to 15 months for all the places of articulation. The negative values were shorter in the child than in adults from 6 to 12 months. The child produces values around +15 ms for the unvoiced occlusives (cf. Figure 3). These values are shorter but not very different from French adults values: around +30 ms (Serniclaes, 1987) and values around -25 ms for voiced occlusives while French adult negative values are around -130ms (Serniclaes, 1987).



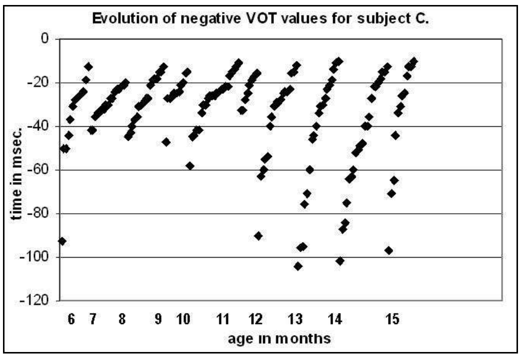


Figure 3 (left): Evolution of positive VOT values between 6 and 15 months for subject C. **Figure 4 (right):** Evolution of negative VOT values

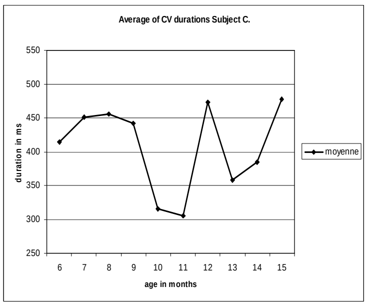
We cannot conclude anything about the oro-laryngeal control from 6 to 12 months even though the values for positive VOT are not very different from the adults' values (Figure 3), since we cannot decide whether the two distinct categories that appear at that age are due to aerodynamic constraints, to a side effect of our measuring method, or to real active control of the coordination. But what is clearer is that at 12 months, the negative values become longer. The average negative value is around -40 ms and child is able to produce some VOT values at -100 ms (cf. Figure 4). Then we can suppose that the child has already acquired the ability to start the vocal folds vibrations after the release of the supra-glottal occlusion at 6 months. But she will be able to begin mastering the vocal folds vibrations during the occlusions at 12 months.

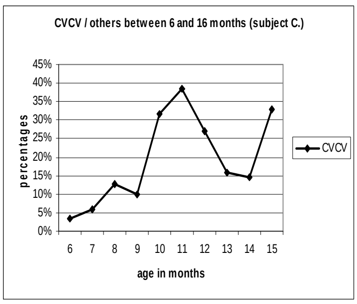
3. The rhythmic control.

What rhythm is needed to make a good first word? First you have to master your mandibular oscillations enough to produce varied rhythms, and not only the default rhythm which is imposed by the biomechanical properties of the mandibular system. Second you have to be able to stop the oscillations after two cycles, since 2/3 of the words in the world's languages are 2 or 3 syllables long (Rousset, 2004). First words are actually in a majority dissyllabic (Gerken, 1994, Boysson-Bardies, 1996). And third, your word will be recognized as a word if you can adapt it to the prosodic pattern of your mother tongue. We investigated these three questions in our data: (i) when does the control of the rhythmic mandibular oscillations appear?; (ii) what is the evolution for CVCV proportions between 6 and 15 months?; and (iii) when does the child adapt the rhythm of his utterances to the iambic pattern of French adult words?

3.1. Getting free from the biological rhythm.

The jaw is a large bone moved by powerful muscles that are inserted into the skull. The whole system is a very heavy one with a lot of inertia. Considering babbling as resulting from the ability to initiate mandibular oscillations does not mean that the baby will be able to produce various rhythmic patterns of oscillations. Rather, it may be supposed that the oscillatory rhythm will tend to centre on the biological rhythm of the system with no voluntary variation allowed. The emergence of rhythmic variations will then stand as evidence that the child is gaining control over the mandibular system. We have first measured the frequency of 2807 CV syllables in disyllabic and plurisyllabic productions from 6 to 15 months. Our results (cf. Figure 5) show that from 6 to 9 months the mean syllable duration is quite stable, around 450 ms. The mandible appears to be moving with a very monotonous rhythm, which may be called the biological rhythm of this motor system. The child then produces shorter syllable durations, around 300 ms, at 10 and 11 months: he seems to be gaining control over the speed of the mandibular system.





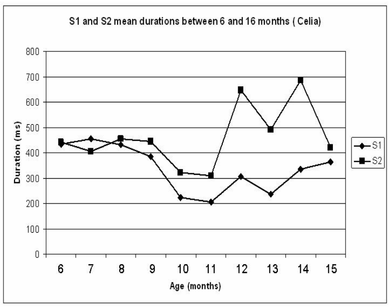


Figure 5 (left): Evolution of the mean CV durations from 6 to 15 months, subject C. **Figure 6** (center): Evolution of CVCV proportions vs. all other productions, 6 to 15 months for subject C. **Figure 7** (right): Evolution of the mean duration of S1 and S2 for subject C, 6 to 15 months.

Rhythmic control is therefore emerging after 9 months, and the faster rhythm of the mandible is not due to the growth of this bone structure, as is shown by models of cranial growth (Boë, 2006, pers.comm.): the mandibular bone undergoes a quite steady growth, as opposed to what is happening for cranial growth, which shows more important jumps. From 11 to 15 months, the rhythm production undergoes different variations. The difference between babbling rhythm and first words rhythm has been investigated and showed no significant variation. These variations in the global rhythm of CV productions are to be explained by more complex evolutions in the control of rhythm, so we investigated the evolution of inter-syllabic sequences and more particularly the nature of the CVCV productions.

3.2. Emergence of a CVCV pattern?

Knowing that (i) children are exposed in majority to words that are composed of 2 or 3 syllables, and that (ii) CVCVs are the most frequent productions present in babbling among all the reduplicated productions of CVs (as shown from a babbling sample from 6 to 18 months by Ducey & al. (2006)), we observed the evolution of CVCV proportions in babbling and first words. The results show (cf. Figure 6) that between 6 and 9 months, there is very few CVCV (between 1% and 10 %) vs. other productions in babbling. At 10 months, the proportions increase dramatically to reach a peak of 40 % of the babbling productions at 11 months. At the age of 13 and 14 months the figures decrease and there is less CVCV than other patterns in the production of the child. We can attribute these results to the fact that this child produces also a lot of monosyllabic utterances as first words at this time. Anyway the most significant outcome is that the appearance of first words seems to be coinciding with a peak in CVCV production.

3.3. Adaptation to the rhythmic prosodic pattern of French words?

French is an oxytonic language, which means that it has iambic disyllabic patterns, with a second syllable that is longer than the first one. To study the adaptation to the rhythmical pattern of French adults' words – Syllable2 (S2) longer than Syllable1 (S1) in a CVCV structure –, we measured the durations of 864 CVCV from 6 to 15 months. Figure 7 shows the evolution of the mean durations of S1 and S2 between 6 and 15 months: S1 and S2 begin with the same duration around 400 ms from 6 to 8 months. S2 becomes slightly longer than S1 at 10 and 11 months, then the difference increases considerably from 12 to 14 months (S2 around 600-700 ms.), and comes down again at 15 months, around 400 ms. Our interpretation is that the adaptation to the French adult-like pattern starts appearing in this child at 10 months, it is very much emphasized from 12 months to 14 months, and the baby comes back to more normal ratios at 15 months. This pattern was also described in a French baby by Konopczynski (1998). We have observed separately the evolution of S1 and S2 durations during babbling productions and first words and we got no difference between babbling and first words for the durations of S1 and S2 from 10 months to 15 months (no first words before 10 months). We also compared from 6 to 15 months the evolution of the duration of the syllable in position S1 and S2 in CVCV, and the duration of all other syllables in more than dissyllabic productions. Results show that the global variation in CV durations is largely explained by the evolution of CVCV structures during that period. At 15 months, the pattern is no longer present but we can suppose that it is due to the acquisition of other controls. First words are therefore rhythmic patterns. 10 months is the time when first words appear in this child, and it is also the time when the child begins to be able to produce CVCV sequences, with an inter-syllabic rhythm that will allow the speakers of his language community to recognize it as a word of the language: he is beginning to adapt himself to the prosodic features of his mother tongue.

4. Conclusion

Studying speech development during the first year of life is a way to bring out a number of constraints and articulatory complexities that need to be mastered, which in turn shed a specific light on the control specificities in adult speech. Our work is guided by such principles and the constraints that come out of our data are introduced into speech models that have been developed for several years from adult language data. The Perception for Action Control theory (PACT – Schwartz & al, 2006) intends to describe language units, from the perspective of sensory-motor interactions and with constant reference to phonology, as resulting from perceptive and motor constraints, their organization and morphogenesis being constrained from the start by the progressive emergence of motor skills during ontogeny. The emergence of phonology, which is the focus of all developmental studies, has to be understood as the structuration of sensory-motor maps built from the interactions between multimodal perception information, the elaboration of specific cognitive representations and the mastering of the degrees of freedom of the vocal tract through the emergence of motor controls sequenced by universal constraints.

Acknowledgements

We are deeply grateful to Stefanie Brosda, Solange Rossato,Virginie Ducey-Kauffman, Romain Trollat, Christian Abry, and to our young subjects and their families.

References

Amelot, A., Rossato, S. (2006), Velar movements for the [±nasal] feature for two French speakers, *ISSP'06*, Ubatuba, Brasil.

Boë, L.-J, Maeda, S. (1998). Modélisation de la croissance du conduit vocal. *Journées d'Études Linguistiques "La Voyelle dans tous ses états". Nantes*, 98-105.

Brosda, S. (1999). *De la variation dans le babillage canonique: l'apprentissage sensori-moteur*, MPhil thesis, ICP, Grenoble.

Davis, B.L. & MacNeilage, P.F. (2000). An embodiment perspective on the acquisition of speech perception, *Phonetica*, 2000, 57, 229-241.

Delattre, P. (1954). Les attributs acoustiques de la nasalité vocalique et consonantique, *Studia Linguistica*, 8, 103-109.

Ducey-Kaufmann V., Abry C., Vilain C. (2006). When the */Speech Frame/* meets the */Sign Frame/* in a developmental framework, submit to ELA2005 Proceedings (Lyon).

Eimas, P. & al. (1971). Speech perception in infants, *Science*, 171, 303-306.

Gerken, L. (1994). A metrical template account of children's weak syllable omission from multisyllabic words, *Journal of Child Language*, 21, 565-584.

Green, J. R., Moore, C. A., Higashikawa, M. & Steeve, R. W. (2000). The physiologic development of speech motor control : lip and jaw coordination, *Journal of Speech, Language and Hearing Research*, 43, 239-255.

Kinney, A. L., & MacNeilage, P.F. (2001). Nasalization in vocalizations of pre-babbling infants: investigating an acoustic correlate, *IVth International Speech Motor Conference, Nijmegen, the Netherlands,* 28, 113-116.

Kinney, A. L. MacNeilage, P.F., Davis, B.L. (2002). The role of nasalization in early vocal output, submitted to *Phonetica*

Konopczynski, G. (1998). Interactive Developmental Intonology (IDI): Theory and application to French, *Revue Parole*, 7-8, 177-201.

Koopmans Van Beinum, F. J. (1999). AMSTIVOC: testing and elaborating the Amsterdam system for transcription of infant vocalizations, *IFA Proceedings*, 23, 91-102.

Lisker, L. & Abramson, A.S. (1964). A cross-language study of voicing in initial stops, Acoustical measurements, *Word*, 20, 384-422.

MacNeilage, P.F. (1998). The frame/content theory of evolution of speech production, *Behavioral and Brain Sciences*, 21, 499-546.

Matyear, C.L., Macneilage, P.F. & Davis, B.L. (1998). Nasalization of vowels in nasal environments in babbling: evidence for frame dominance, *Phonetica*, 55, 1-17.

Munhall, K.G. & Jones, J.A. (1998). Articulatory evidence for syllabic structure. *Behavioral and Brain Sciences*, 21(4), 524-525.

Rossato, S., Badin, P., Bouaouni, F. (2003). Velar movements in French: an articulatory and acoustical analysis of coarticulation, *XVth ICPhS*, Barcelona, 3141-3144.

Rousset, I. (2004). *Structures syllabiques et lexicales des langues du monde: données, typologies, tendances universelles et contraintes substantielles,* PhD thesis, ICP, Grenoble.

Schwartz, J.L. & al. (1997) The Dispersion-Focalization Theory of vowel systems, *Journal of Phonetics*, 25, 255-286.

Schwartz, J.L., Boë, L.J., Abry, C. (2006). Linking the Dispersion-Focalization Theory and the Maximum Utilization of Available Distinctive Features principle in a Perception-for-Action-Control Theory. In *Festschrift to John Ohala*(in press).

Serniclaes, W. (1987). *Etude expérimentale de la perception du trait de voisement des occlusives du Français,* Thèse, Institut de phonétique, Université Libre de Bruxelles.

Stefanuto, M., Vallée, N. (1999). Consonant systems: From universal trends to ontogenesis, *XIVth ICPhS,* San Francisco, 1973-1976.

Sundberg U. & Lacerda, F. (1999). Voice onset time in speech to infants and adults, *Phonetica*, 56, 186-199.

Tubach, J.P. & Boë, L.J. (1990). *Un corpus de transcription phonétique (300.000 phones) : constitution et exploitation statistique*, ENST, Paris.

Vallée, N. (1994). *Systèmes vocaliques: de la typologie aux prédictions,* PhD thesis, Grenoble.

Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., & Miller, J. (1985). From babbling to speech: A re-assessment to the continuity issue. *Language*, 61, 397-445.

Vilain, A., Abry, C., Badin, P., Brosda, S. (1999). From idiosyncratic pure frames to variegated babbling: Evidence from articulatory modelling, *ICPhS'99,* San Francisco, 3, 2497.